VERSION WITH MARKINGS TO SHOW CHANGES MADE

IN THE SPECIFICATION:

The paragraph beginning at page 4, line 4, has been amended as follows:

[OBJECTS AND ADVANTAGES] SUMMARY OF THE INVENTION

The paragraph beginning at page 4, line 5, has been amended as follows:

[Accordingly, it is a primary object] Various aspects of the present invention [to] provide [a method] systems and methods for ranking documents in a linked database. [It] One aspect [is another object of the invention to provide such a method that] provides an objective ranking based on the relationship between documents. Another [object] aspect of the invention is [to provide] directed to a technique for ranking documents within a database whose content has a large variation in quality and importance. Another [object] aspect of the present invention is to provide a document ranking method that is scalable and can be applied to extremely large databases such as the world wide web. Additional [objects and advantages] aspects of the invention will become apparent in view of the following description and associated figures.

The paragraph at page 4, line 18, has been deleted.

The paragraph beginning at page 4, line 19, has been amended as follows:

[The] One aspect of the present invention [achieves the above objects by] is directed to taking advantage of the linked structure of a database to assign a rank to each document in the database, where the document rank is a measure of the importance of a

document. Rather than determining relevance only from the intrinsic content of a document, or from the anchor text of backlinks to the document, [the] a [present] method consistent with the invention determines importance from the extrinsic relationships between documents. Intuitively, a document should be important (regardless of its content) if it is highly cited by other documents. Not all citations, however, are necessarily of equal significance. A citation from an important document is more important than a citation from a relatively unimportant document. Thus, the importance of a page, and hence the rank assigned to it, should depend not just on the number of citations it has, but on the importance of the citing documents as well. This implies a recursive definition of rank: the rank of a document is a function of the ranks of the documents which cite it. The ranks of documents may be calculated by an iterative procedure on a linked database.

The paragraph beginning on page 5, line 19, has been amended as follows:

In one aspect of the invention, a computer implemented method is provided for [calculating an importance rank for N linked nodes] scoring [of a linked database] linked documents. The method [comprises the steps of:] includes identifying a plurality of documents, at least some of the documents being linked documents, at least some of the documents being linking documents, and at least some of the documents being both linked documents and linking documents, each of the linked documents being pointed to by a link in one or more of the linking documents; assigning a score to each of the linked documents based on scores associated with the one or more linking documents; and processing the linked documents according to the assigned scores.

In accordance with another implementation consistent with the present invention, a method for scoring linked documents includes receiving a search query from a user; identifying a plurality of documents responsive to the search query; locating incoming links to the identified documents from corresponding linking documents; assigning a score to each of the identified documents based on a number of the incoming links to the identified document and an importance of the incoming links; creating a ranked list based on the scores of the identified documents; and presenting to the user information about the identified documents in an order that is based on the ranked list.

In accordance with yet another implementation consistent with the present invention, a method for organizing linked nodes includes determining first link information for a linked node; determining second link information for linking nodes that link to the linked node; and calculating a score for the linked node based on both the first link information and the second link information.

In accordance with a further implementation consistent with the present invention, a method of organizing linked documents includes: (a) identifying a first linked document; (b) identifying links between linking documents and the first linked document; (c) assigning a weight to each of the identified links; (d) determining a score for the first linked document based on (i) a number of the identified links between the linking documents and the first linked document, and (ii) the weights assigned to each of the identified links; (e) repeating steps (a) – (d) for a second linked document; and (f) organizing the first and second linked documents based on the determined scores.

In accordance with another implementation consistent with the present invention, a method of organizing documents includes identifying a plurality of linked documents;

identifying primary linking documents that link to the linked documents; identifying secondary linking documents that link to the primary linking documents; assigning a score to each of the linked documents based on (i) the number of links between the primary linking documents and the linked document and (ii) the number of links between the primary linking documents and the secondary linking documents; and organizing the linked documents according to the assigned scores.

Additional aspects, applications and advantages will become apparent in view of the following description and associated figures.

[(a) selecting an initial N-dimensional vector $\mathbf{p}_0$;

(b) computing an approximation $\mathbf{p}_n$ to a steady-state probability $\mathbf{p}_\infty$ in accordance with the equation $\mathbf{p}_n = \mathbf{A}^n\mathbf{p}_0$, where **A** is an NxN transition probability matrix having elements **A**[i][j] representing a probability of moving from node i to node j; and

(c) determining a rank r[k] for a node k from a k^{th} component of $\mathbf{p}_n$.]

The paragraph beginning at page 5, line 31, has been deleted.

The paragraph beginning at page 6, line 19, has been amended as follows:

Although the following detailed description contains many specifics for the purposes of illustration, anyone of ordinary skill in the art will appreciate that many variations and alterations to the following details are within the scope of the invention. Accordingly, the following [preferred embodiment] embodiments of the invention [is] are set forth without any loss of generality to, and without imposing limitations upon, the claimed invention. For support in reducing the present invention to practice, the inventor

acknowledges Sergey Brin, Scott Hassan, Rajeev Motwani, Alan Steremberg, and Terry Winograd.

The paragraph beginning at page 11, line 27, has been amended as follows:

In [a preferred] one particular embodiment, a finite number of iterations are performed to approximate $\mathbf{p}_\infty$. The initial distribution can be selected to be uniform or non-uniform. A uniform distribution would set each component of $\mathbf{p}_0$ equal to 1/N. A non-uniform distribution, for example, can divide the initial probability among a few nodes which are known a priori to have relatively large importance. This non-uniform distribution decreases the number of iterations required to obtain a close approximation to $\mathbf{p}_\infty$ and also is one way to reduce the effect of artificially inflating relevance by adding unrelated terms.

The paragraph beginning at page 12, line 6, has been amended as follows:

In [a preferred] another particular embodiment, the transition matrix **A** is given by

$\mathbf{A} = \alpha/N\mathbb{1} + (1-\alpha)\mathbf{B}$,

where $\mathbb{1}$ is an NxN matrix consisting of all 1s, α is the probability that a surfer will jump randomly to any one of the N nodes, and **B** is a matrix whose elements **B**[i][j] are given by

$$B[i][j] = \begin{cases} 1/n_i & \text{if node i points to node j} \\ 0 & \text{otherwise} \end{cases},$$

where n_i is the total number of forward links from node i. The $(1-\alpha)$ factor acts as a damping factor that limits the extent to which a document's rank can be inherited by children documents. This models the fact that users typically jump to a different place in the web after following a few links. The value of α is typically around 15%. Including this damping is important when many iterations are used to calculate the rank so that there is no artificial concentration of rank importance within loops of the web. Alternatively, one may set $\alpha=0$ and only iterate a few times in the calculation.

The paragraph beginning at page 12, line 27, has been amended as follows:

[There] Consistent with the present invention, there are several ways that this method can be adapted or altered for various purposes. As already mentioned above, rather than including the random linking probability α equally among all nodes, it can be divided in various ways among all the sites by changing the $\mathbf{1}$ matrix to another matrix. For example, it could be distributed so that a random jump takes the surfer to one of a few nodes that have a high importance, and will not take the surfer to any of the other nodes. This can be very effective in preventing deceptively tagged documents from receiving artificially inflated relevance. Alternatively, the random linking probability could be distributed so that random jumps do not happen from high importance nodes, and only happen from other nodes. This distribution would model a surfer who is more likely to make random jumps from unimportant sites and follow forward links from important sites. A modification to avoid drawing unwarranted attention to pages with artificially inflated relevance is to ignore local links between documents and only consider links between separate domains. Because the links from other sites to the

document are not directly under the control of a typical web site designer, it is then difficult for the designer to artificially inflate the ranking. A simpler approach is to weight links from pages contained on the same web server less than links from other servers. Also, in addition to servers, internet domains and any general measure of the distance between links could be used to determine such a weighting.

The paragraph beginning at page 14, line 8, has been amended as follows:

[The present method has] Various implementations of the invention have the advantage that the convergence is very fast (a few hours using current processors) and it is much less expensive than building a full-text index. This speed allows the ranking to be customized or personalized for specific users. For example, a user's home page and/or bookmarks can be given a large initial importance, and/or a high probability of a random jump returning to it. This high rating essentially indicates to the system that the person's homepage and/or bookmarks does indeed contain subjects of importance that should be highly ranked. This procedure essentially trains the system to recognize pages related to the person's interests.

The paragraph beginning at page 15, line 10, has been amended as follows:

[Perhaps the most] Another important application and embodiment of the present [ranking technique is to enhance] invention is directed to enhancing the quality of results from web search engines. In this application of the present invention, [the] a ranking method [of] according to the invention is integrated into a web search engine to produce results far superior to existing methods in quality and performance. A search engine

employing [the] a ranking method of the present invention [has all the advantages of] provides automation while producing results comparable to a human maintained categorized system. In this approach, a web crawler explores the web and creates an index of the web content, as well as a directed graph of nodes corresponding to the structure of hyperlinks. The nodes of the graph (i.e., pages of the web) are then ranked according to importance [according to the method] as described above in connection with various exemplary embodiments of the present invention.

The paragraph beginning at page 15, line 25, has been amended as follows:

The search engine is used to locate documents that match the specified search criteria, either by searching full text, or by searching titles only. In addition, the search can include the anchor text associated with backlinks to the page. This [idea] approach has several advantages in this context. First, anchors often provide more accurate descriptions of web pages than the pages themselves. Second, anchors may exist for images, programs, and other objects that cannot be indexed by a text-based search engine. This also makes it possible to return web pages which have not actually been crawled. In addition, the engine can compare the search terms with a list of its backlink document titles. Thus, even though the text of the document itself may not match the search terms, if the document is cited by documents whose titles or backlink anchor text match the search terms, the document will be considered a match. In addition to or instead of the anchor text, the text in the immediate vicinity of the backlink anchor text can also be compared to the search terms in order to improve the search.

The paragraph beginning at page 16, line 10, has been amended as follows:

Once a set of documents is identified that match the search terms, the list of documents is then sorted with high ranking documents first and low ranking documents last. The ranking in this case is [defined as] a function which combines all of the above factors such as the objective ranking and textual matching. If desired, the results can be grouped by category or site as well.

The paragraph beginning at page 16, line 18, has been amended as follows:

It will be clear to one skilled in the art that the above [embodiment] <u>embodiments</u> may be altered in many ways without departing from the scope of the invention. Accordingly, the scope of the invention should be determined by the following claims and their legal equivalents.